Author's response to reviews

Title: Psychometric Properties of the AHRQ Hospital Survey on Patient Safety Culture

Authors:

   Joann S Sorra PhD (joannsorra@westat.com)
   Naomi Dyer PhD (naomidyser@westat.com)

Version: 3 Date: 13 May 2010

Author's response to reviews: see over
DATE: May 12, 2010

TO: Miss Gabriella Anderson & Reviewers Marleen Smits and Cordula Wagner, BMC Health Services Research

SUBJECT: Research Article on the psychometric properties of a survey

Article Title: Manuscript 7542290873507929--Psychometric Properties of the AHRQ Hospital Survey on Patient Safety Culture

Authors: Joann S. Sorra, Ph.D. and Naomi Dyer, Ph.D., Westat

This cover letter accompanies our revised manuscript which we are submitting for publication in BMC Health Services Research. We are grateful to the two reviewers for their thoughtful comments and have included a point-by-point response, indicating where changes were made to the manuscript. Some changes are shown with tracked changes, others are highlighted in the manuscript.

Author information is below:

Corresponding author:
Joann S. Sorra, Ph.D.
Senior Study Director
Westat
1600 Research Blvd.
Rockville, MD 20850
phone: 301-294-3933
fax: 301-315-5912

Naomi Dyer, Ph.D.
Westat
1600 Research Blvd.
Rockville, MD 20850
phone: 301-610-8842
fax: 301-315-5912

Responses to Reviewer 1:

1) In the results section, there is a lot of text describing steps that were taken in the analyses, reasons why these steps were taken and criteria that were used for interpretation. These text parts should be placed in the methods section. It mainly concerns the following parts:
   • Page 9 “Individual Level Factor Analysis”: “Factor analyses were conducted to confirm whether……should be .40 or greater” (6 lines)
   • Page 9 “Individual Level Factor Analysis”: “Another statistic examined to determine the adequacy of a factor……at least 50% of the variance should be accounted for by the composite” (4 lines)
• Page 9/10 “Multilevel Analyses”: “Individuals responding to the Hospital SOPS are located....should be conducted to account for the multilevel nature of the data” (18 lines).
• Page 11 “Multilevel Confirmative Factor Analysis (MCFA)”: “MCFAs were conducted using MPlus Version 5.1....a value less than 0.08 is considered a good fit” (24 lines)
• Page 13: “Reliability Analysis”: “Internal consistency reliability was examined by calculating.....is a alpha of at least .70” (5 lines)
• Page 14: “Interrelations Among the 12 Patient Safety Culture Composites”: “Intercorrelation among the patient safety culture composites were explored at three levels..... and avoid problems with multicollinearity”. (8 lines)

RESPONSE: As recommended, we moved the criteria for interpretation of the results from the Results section to the Analyses section. In doing this, we added sub-headers for readability based on each type of analysis. See revised Analyses section of the paper starting on page 8.

2) The authors use criteria that hospitals and units have to meet to be kept in the database (page 6, “Analysis Dataset”). One is that a hospital has to have more than one unit that responded to the questionnaire. Why is the cut-off point between one and two units? Is two units enough for the analyses?

RESPONSE: The rationale for more than one unit was that if a hospital only had one unit, the unit level analyses would be equivalent to the hospital level analyses. Furthermore, hospitals that have more than one unit are more representative of a hospital and therefore we opted to restrict the data to those with more than one unit. Finally, we wanted to be as inclusive as possible, so did not set a higher criteria (e.g., at least 3 units) because some of the hospitals are small (6-24 beds) or did not sample across all units.

3) Another criterion is that units have to have 3 respondents or more. This is a very low number, especially in larger units with for example more than 30 employees. Shouldn’t there be different cut-off points for large and small units? For large units, a cut-off point of 10 is more agreeable (then you really have a GROUP of people responding to this group culture survey).

RESPONSE: We set the criteria for number of respondents within a unit to be 3 or more because some of the hospitals are very small (6-24 beds) and may only have 3-5 individuals working within a unit. If we had set a higher criterion, we would have virtually eliminated this type of small hospital from our analyses. We opted to be most inclusive for these analyses. As a note, there can be as few as 2 respondents within a group to conduct multilevel analyses.

4) More information on the response rate calculation is needed. Was it related to only the participating units of the hospitals or the whole hospital? Was the whole unit invited
to participate in the survey (so that if the response rate was 100%, every employee of the unit responded) or was a sample taken?

**RESPONSE:** We have added two sentences in the Sample and Response Rate section on page 6 to provide more detail. These sentences are also provided here:

“Response rates were calculated from self-reported numbers provided by the hospitals indicating how many staff were asked to participate in the survey across the hospital. The remaining hospitals surveyed a combination of selected staff and/or selected departments. Approximately 77% of the hospitals indicated they surveyed all staff, or a sample of all staff, from all departments.”

5) Page 7 “Measures”: please provide more information on the analyses that were performed previously to define the 12-composite-structure. This composite-structure is now described as a given fact, but, as there is no previous paper of the authors that describes the exploratory analyses, more information should be given in this paper.

**RESPONSE:** Because we introduce the AHRQ survey under the first section of the Methods section “Development of the Survey and Comparative Database,” we have added some more description of the analyses cited in the technical report in that section, indicating that exploratory and confirmatory factor analyses were conducted, the goals of the psychometric analysis, and reliabilities.

6) The confirmative factor analyses were performed for each a priori patient safety composite separately. There is a method to place all composites together in a model to confirm the factor structure. Did the authors consider this structural equation modeling (SEM) technique? (for example with LISREL)

**RESPONSE:** We initially attempted to perform the multilevel confirmatory factor analysis (MCFA) on the entire model as suggested. However, given the complexity of our model with 12 composites with 42 items there were issues with the multilevel model (taking into account individual, unit, and hospital-level variance simultaneously) failing to converge using the MPlus software. We are not aware of any published articles that show MCFAs conducted on as many as 12 composites simultaneously. Several articles in which MCFAs were performed only examined 1-3 composites at time (c.f., Reise, S.P., Ventura, J., Nuechterlein, K.H., & Kim, K.H., 2005; Zhang, N.J, & Wan, T.H., 2005).

7) ICCs are calculated per item. Please add the composite ICCs.

**RESPONSE:** We used Mplus software to calculate the ICCs for this data. Mplus provides item ICCs does not provide composite or latent variable ICC results. We are unaware of a method for calculating composite ICCs that takes into account the multilevel nature of the data.
8) The significance of the ICCs is not tested. Why not?

**RESPONSE:** We used Mplus software to calculate the ICCs for this data. Mplus provides the item ICCs but does not provide statistical significance for ICCs. Muthen (who developed the Mplus software) suggests that design effects are a better indicator to examine rather than ICCs. We provide ICCs for those who are more familiar with that statistic, but also present results for design effects (which do not have statistical significance tests).

9) My colleagues and I have written a paper on a study we performed to examine the multilevel structure of the HSOPS with the Dutch version of the HSOPS. The authors have referred to our paper in a short list of different countries that have examined psychometric properties (reference nr 38). However, I miss a comparison of the results of our multilevel study with the current study in the discussion section: a comparison on ICC scores at the different levels and intercorrelations of the composites.

**RESPONSE:** Because we were examining the measurement model and displaying ICCs at the item level and the Smits et al (2009) paper examines the latent model and displays ICCS at the composite level, we do not think a direct comparison of ICCs would be appropriate.

**Minor Essential Revisions**
10) Page 17: the authors list methodological approaches that can be used to identify patient safety vulnerabilities. They should add a method that is commonly used in many countries over the world: patient record review (for example record review based on the Harvard Medical Practice Study).

**RESPONSE:** We have added reference to medical record review.

11) Table 6: First row, first column: Please replace “Hospital Survey Dimensions and Items” by “Hospital Survey Composites”.

**RESPONSE:** We have made this correction so it now reads “Hospital Survey Composites”.

**Discretionary Revisions**
12) Tables: There are a lot of tables in the paper. Table 1 is the least important for a paper on the psychometric properties of a questionnaire and can be deleted.

**RESPONSE:** Table 1 shows the distribution of hospitals by bed size and provides information about the representativeness of the hospitals in the data set. We think it will be of interest to readers, but the editors may remove this if they do not think it is important.
13) Page 13: “Reliability analysis”: Please delete “on the row next to the composite title”. This is evident.

   RESPONSE: We have made this correction.

14) Discussion: The authors can add another strength of the survey: The amount of clustering at higher levels indicates that the survey measures what it is supposed to: group culture (and not only individual attitudes).

   RESPONSE: We have added a sentence about this in the first paragraph of the discussion.


   RESPONSE: Thank you. We have added this citation.

Responses to Reviewer 2:

Minor Revisions

1. The authors describe that the average response rate for the 331 hospitals was 55% (range: 6% to 100%), with an average of 289 respondents per hospital (range: 11 to 3,684). Is it valid to include a hospital with a response rate of 6% or only 11 respondents? This is a very small number.

   RESPONSE: We conducted independent analyses on the 2009 Hospital Survey on Patient Safety Culture Database to examine if there should be a minimum response rate for hospitals submitting to the database. We found that there was no clear evidence for implementing a response rate cut-off criterion. When comparing no response rate cutoff (including all hospitals regardless of response rate) to a response rate cutoff as high as 50%, average safety culture composite scores for the hospitals only changed by 1 percentage point. Therefore we decided against setting a response rate criterion since including results from hospitals with lower response rates did not seem to affect the overall data.

2. There is some repetitive information, e.g. that the survey consists of 42 items and 12 dimensions results.
RESPONSE: We have deleted repetitive mentions and now have only 3 occurrences of this phrasing (abstract, measures, and discussion).

3. The authors describe that the chi-square tests in table 6 for six composites were significant, whereas a non-significant chi-square indicates good fit. So, the obvious conclusion is that there is no good fit. The authors explain then that in large samples, as they have, the tests will frequently be significant. Why using this test in the first place and not the others more appropriate tests for the sample size.

RESPONSE: In a review of reporting practices for CFAs, it was found that over 89% of articles report the chi-square value (Jackson, Gillapsy, and Purc-Stephenson, 2009). We included the chi-square test, despite its known limitations, as it is one of standard metrics reported in confirmatory factor analyses (CFA) and we believed some researchers will still want to see the results for it.

4. For this study it is not important to know in detail non significant correlations between 'number of events reported' and the 12 composites.

RESPONSE: We have adjusted that section to only detail the significant correlations.

Conclusion

5. The staffing composite fell below various cut-off points. I understand why the authors want to maintain that composite, but why not improving the questions based on the written comments. Maybe there are other issues of staffing that are more important or better in creating one construct.

RESPONSE: The goal of the present paper was to simply present psychometric results. However, in a “Version 2.0” of the AHRQ Hospital survey, we will absolutely consider ways to revise the staffing questions and composite to improve it.

References
